The comprehensive midtrimester test: High-sensitivity Down syndrome test

Ray Bahado-Singh, MD,a Shohreh Shahabi, MD,a Mehmet Karaca, MD,a Maurice J. Mahoney, MD,a Laurence Cole, PhD,b and Utku A. Oz, MDc

New Haven, Conn, Albuquerque, NM, and Mersin, Turkey

OBJECTIVE: The purpose of this study was to develop a highly sensitive algorithm for midtrimester Down syndrome detection.

STUDY DESIGN: Urine (hyperglycosylated human chorionic gonadotropin, β-core fragment of human chorionic gonadotropin), serum (α-fetoprotein, human chorionic gonadotropin and unconjugated estriol [uE3]), and ultrasound biometry (nuchal thickness, humerus length, the presence of gross ultrasonographic anomalies), and maternal age were measured at genetic amniocentesis. Stepwise logistic regression analysis was used to identify the most significant markers. A multivariate Gaussian algorithm plus age was used to derive patient-specific Down syndrome risk. Sensitivity and false-positive rates at different risk thresholds and the area under the operating characteristic curve were determined. A probability value of <.05 was significant.

RESULTS: There were 568 study cases with 17 Down syndrome cases (3.0%). The mean (±SD) maternal and gestational ages for the study group were 36.9 (±3.5) years and 16.2 (±1.4) weeks, respectively. The significant markers were nuchal thickness (P = .0001), hyperglycosylated human chorionic gonadotropin (P < .001), and β-core fragment (P < .002). Neither maternal age nor gross sonographic anomaly contributed significantly to Down syndrome detection. The comprehensive midtrimester test was extremely efficient for Down syndrome detection in advanced maternal age only cases with a sensitivity of 92.3% at a 0.8% false-positive rate. In women <35 years old, all the Down syndrome cases were detected at 2.2% false-positive rate. For the overall population, the sensitivity was 93.7% at 5% false-positive rate.

CONCLUSION: In a preliminary study, the comprehensive midtrimester test appeared highly sensitive in different age groups. Gross anomaly detection was not required for high performance, which makes the comprehensive midtrimester test potentially suitable for low-risk screening and as an alternative to amniocentesis in women who wish to avoid the procedure. This was a small study; thus, the clinical value of this test can only be established in large trials. (Am J Obstet Gynecol 2002;186:803-8.)

Key words: Down syndrome, urine, screening
In a series of articles, we have reported that a urine mid-trimester marker (hyperglycosylated hCG [HhCG], also called invasive trophoblast antigen) has the highest Down syndrome screening performance reported for a single marker in pregnancy. The reported detection rate is 80%, at a 5% false-positive rate. In addition, urine \( \beta \)-core fragment, a breakdown product of hCG, when measured in fresh maternal urine, detects 60% of Down syndrome cases at a 5% false-positive rate.

The purpose of the study was to develop a comprehensive mid-trimester test (CMT), based on the evaluation of multiple urine, serum, and ultrasound markers for each patient. We were interested in developing a test with high Down syndrome screening performance but without the concerns associated with combined first- and second-trimester screening. In addition, we wanted an algorithm that required a minimum number of markers. Potential benefits of the latter include reduced cost and enhanced practicability in the deployment of the test in clinical settings.

Methods

After Institutional Review Board (Yale University School of Medicine) approval was obtained, patients who were scheduled for midtrimester genetic amniocentesis gave verbal consent and volunteered a spot urine specimen for the measurement of urine biochemical analytes. All the patients used in this study had blood drawn for a triple screen before amniocentesis or had a serum specimen drawn at the time of the amniocentesis for AFP measurement (for the prediction of perinatal complications). Subsequently, unconjugated estriol and hCG levels were measured from the remaining serum in the AFP-only group. Nuchal thickness and humerus length measurements and a targeted ultrasound scan for gross morphologic defects were performed before the genetic amniocentesis. The technique for measurement of urine \( \beta \)-core fragment8,10 and HhCG6,7 has been extensively described. Urine marker levels were standardized to creatinine level. Values were expressed as multiples of the gestational age-based medians, as described previously.6,9

The urine, serum, and ultrasound markers were measured without knowledge of the karyotype. Urine \( \beta \)-core fragment was measured in fresh urine; HhCG was measured after 1 freeze-thaw cycle after storage at \(-20^\circ\)C. The different markers were compared to determine which combination was the best midtrimester predictor of fetal Down syndrome.

When multiple urine markers, \( \beta \)-core fragment and total estriol11 or urine combined with serum markers (namely, \( \beta \)-core fragment), total estriol and serum AFP14 or ultrasound biometry plus urine \( \beta \)-core fragment, and total urine estriol15-17 are used, the Down syndrome screening performance was significantly increased15-17 and exceeded the performance of the currently available midtrimester algorithms. The patients in this study represented a subgroup drawn from the previous studies.6,9,11-17 Only patients from whom all the ultrasound, serum, and urine markers were obtained were included in this study.

The purpose of the current study was to identify the optimal marker combination by direct comparison (all markers obtained from each patient) with a total of 9 serum, ultrasonographic, and urine markers. This has not been done in any previous study by our group or others.

The nuchal thickness was measured in the plane required to obtain the transcerebellar diameter,18 from the outer edge of the occipital bone to the skin surface as close to the midline as possible. This must have been reported to within 1 decimal place (eg, 3.2 mm). Nuchal thickness reported as “normal” (ie, \(<6\) mm) or not reported to within 1 decimal place (eg, 3 mm) was not considered acceptable for the study. The humerus length was measured from 1 end of the shaft of the bone to the other. Similar to the biochemical markers, nuchal thickness19 and humerus length20 measurements were expressed as multiples of the median (MoM) that were standardized for gestational age. As with biochemical markers, there are a number of advantages to expressing biometry measurements as MoM for gestational age. The use of MoM controls for the significant variation in marker values due to gestational age. In addition, the use of MoM values allows the mathematic integration of different categories of markers (ultrasound and biochemical) to derive a single Down syndrome risk estimate, thus making the algorithms more useful clinically.

A detailed anatomic survey was performed in each case by both a sonographer and at least one sonologist. Gross anomalies were defined as structural or morphologic defects or alterations. They include changes in fetal structure, such as cardiac defects or hydrocephalus. This is in contrast with changes in ultrasonic appearance, such as increased bowel echogenicity (subtle markers) or measurement (biometry markers), without a fundamental disruption of the normal fetal anatomy.

Patients were sequentially studied. The exclusionary criterion was the failure to obtain any of the following urine, serum, or ultrasound markers: HhCG, \( \beta \)-core fragment, AFP, hCG, uE3, nuchal thickness, humerus length, and targeted ultrasound evaluation for gross anatomic defects. All patients from whom the 9 markers were obtained and who were 14 to 24 weeks of gestation were included in the study.
Stepwise logistic regression analysis was used to determine the optimal combination of markers for Down syndrome detection. The markers that did not significantly improve screening performance beyond this optimal group were excluded from further consideration. Thus, the most efficient (maximum sensitivity and minimum false-positive rates) and least cumbersome algorithm was derived.

Currently, in multiple serum marker screening, log Gaussian curves of marker distributions for Down syndrome and normal populations are developed for each of the significant markers in the algorithm. The likelihood ratio or odds of carrying a Down syndrome fetus is determined by the ratio of the height of the Gaussian curve in the affected and normal populations at any given MoM value. In the simplest example, single-marker screening (eg, AFP), the likelihood ratio at a given AFP value (MoM) is multiplied by the age-related Down syndrome risk, based on maternal age and the single biochemical marker.\(^{21}\) In the case of multiple analyte screening, an overall or adjusted likelihood ratio that is based on overlapping Gaussian curves can be derived with computerized “matrix” algebra calculations. This online calculation takes into consideration the mean and distributions of marker values in both the normal and affected populations and the degree of overlap or correlation between pairs of analytes in both the Down syndrome and normal groups. Multiple Gaussian curves for individual analytes are thus integrated by converting them into a single “composite” Gaussian distribution.\(^{22}\) A summary likelihood ratio or odds for Down syndrome can be calculated by the combining of the values for each individual analyte. To calculate individual risk, the summary likelihood ratio is multiplied by age-related risk, which was also done in this study. Such calculations constitute the current standard for Down syndrome risk estimation with the use of biochemical analytes.

Based on the Gaussian multivariate algorithm, a numeric Down syndrome risk was calculated for each study patient. Different Down syndrome risk thresholds were used as screening tests, and the corresponding sensitivity and false-positive rates were determined. A receiver operating characteristic (ROC) curve was plotted, with sensitivity against a false-positive rate. The area under the ROC curve and the probability value were measures of the statistical significance of the algorithm. A probability value of <.05 was significant.

**Results**

There were 37 eligible Down syndrome and 1190 normal cases (14-24 weeks of gestation) in which comprehensive ultrasound and urine HhCG and β-core fragment were measured. Of these, 568 cases met the inclusion criteria and constituted the study subjects. The other patients did not have midtrimester triple an-

alyte data and were therefore excluded. Seventeen of the study cases (3.0%) had trisomy 21. The mean (±SD) gestational age of the study population was 16.2 ± 1.4 weeks at amniocentesis. The mean maternal age was 36.9 ± 3.5 years. The indications for karyotype were advanced maternal age (AMA, 78.8%), AMA with abnormal triple screen (3.6%), AMA plus other indications (eg, previous chromosome abnormality and balanced translocation rule out fetal anomaly, 1.5%), abnormal triple screen (8.4%), rule out anomaly (0.4%), and other indications (low AFP, family history, previous pregnancy history, parental anxiety, combination of indications, 7.3%).

There were 6 Down syndrome cases with gross anomalies (4 cardiac defects), 1 omphalocele, and 1 intracranial defect. The following markers were evaluated: serum (AFP, hCG, uE3), urine (HhCG, β-core fragment), and ultrasonographic (nuchal thickness, humerus length, gross anatomic defect).

Stepwise logistic regression indicated that only 3 markers (nuchal thickness, urine HhCG, and β-core fragment of hCG) were significant predictors of Down syndrome (Table I). Neither maternal age, gross anomaly, humerus length, nor any of the serum markers achieved significance in the regression equation. The CMT therefore consisted of urine HhCG, β-core fragment, and nuchal thickness. A priori risk, based on maternal age, was used with the 3 markers to facilitate the calculation of individual Down syndrome risk. For the overall study population, the sensitivities were 85.6% and 99.7% at 1% and 5% false-positive rates. The positive predictive value of the CMT was 68.2%, although the negative predictive value was 99.7% at a threshold risk of >1 of 20 (the threshold at which 85.6% sensitivity and 1.0% false-positive rate is obtained). Adjusted for a Down syndrome prevalence of 1 in 700 (low-risk population), the positive predictive value was 3.3%.

There were 39 cases in which the maternal age was <35 years. Among these, there were 4 cases (4.4%) of Down syndrome. The failure of maternal age to achieve significance in the regression analysis indicates that the performance of the CMT should be similar in different maternal age categories. We therefore assessed the diagnostic accuracy of the CMT in women <35 years old. All Down syndrome cases were identified with a false-positive rate of 2.2%.

The area (±SE) under the ROC curve for the overall study group was 0.990 ± 0.008 (P <.0001, Fig 1). Corre-
The CMT was found to be a highly sensitive and specific test for Down syndrome detection. Table III compares the sensitivity and false-positive rates of the CMT to algorithms that are currently available or have been reported. The diagnostic performance of the CMT is equivalent to that modeled for the integrated test. It should be borne in mind that this study was performed in a high-risk population who underwent amniocentesis. In such populations, along with the high sensitivity achieved, conventional markers also have a high false-positive rate (eg, 20% in the case of the triple test). The CMT, however, achieved high detection rates with low false-positive rates in a high-risk population. In patients who underwent amniocentesis for AMA only, a 92.3% sensitivity at a 0.8% false-positive rate was achieved. This raises the possibility that the CMT might be a reasonable alternative to genetic amniocentesis in this group.

Several features of the CMT deserve specific comment. Maternal age did not contribute significantly to the detection capabilities of the algorithm. The implication of this is that the test can perform equally well in different age groups. This is an important consideration because most Down syndrome fetuses occur in the population of women who are <35 years old. We tested the performance of the CMT in the <35 years group and found it to be an efficient performer, as evidenced by the highly significant area under the ROC curve. This suggests that the test could be suitable for the low-risk population. Another feature of the CMT was interesting in this regard: the detection of gross ultrasound anomalies was not required to achieve the high diagnostic accuracy that was observed. The variability in individual diagnostic acumen, ultrasound equipment, and the significant time commitment required for a detailed anatomy survey disqualifies the latter from being a practical Down syndrome "screening" tool in low-risk populations. The only ultrasound measurement required for the CMT is nuchal thickness, which is a relatively simple and straightforward measurement. The reported success of first-trimester nuchal translucency measurements indicates that simple biometry markers can be used very effectively in population screening (with proper standardization, training of personnel, and ongoing audits).

Another advantage of the CMT is that it avoids the need for venipuncture and its associated risks and expense. Pregnant women routinely provide a urine specimen at each office visit in clinical practice. Urine testing could therefore be performed without significant additional imposition on the patient. To achieve high Down syndrome sensitivity and uniform results over time, β-core fragment hCG assays must use fresh (≤3 days) unfrozen urine specimens. Failure to adhere to these guidelines will result in low sensitivity and variability of results. Similarly, it has been shown that multiple freezing and thawing of urine specimens will result in a marked (approximately 50%) reduction of immunoreactivity of HCG and substantially lower the Down syndrome detection rates.

The failure of the traditional markers (serum AFP, hCG, and uE₃) to achieve significance in the algorithm is not surprising. We had previously demonstrated that fresh urine β-core fragment by itself had a screening performance equal to the triple test, while urine HCG by itself had a statistically superior performance to the serum triple test. Failure of the traditional markers to achieve significance do not negate their well-recognized...
association with Down syndrome. Rather, it reflects the superiority of the 2 urine markers and nuchal thickness. The addition of the serum triple-screen markers would therefore substantially raise the false-positive rate without further significant increase in Down syndrome detection rate in this algorithm.

We have also previously demonstrated the independence of hCG and HhCG. Each has a different cell of origin and substantially different molecular weights and biologic activity. Urine β-core fragment (a breakdown product of hCG) is also substantially independent of HhCG which permits the combination of the 2 markers to improve detection rates.

Although the division of pregnancy into trimesters is deeply entrenched in medical practice, it has no strict biologic basis. Studies that investigate the relative performance of first-trimester and midtrimester markers and the extent of their interaction in the same patient are desirable. The clinical introduction of a combined first- and second-trimester test faces more significant hurdles, however. Concerns regarding wait time for test completion, which results in patient anxiety, and obstetricians’ concerns about withholding abnormal first trimester results from patients must be addressed. In addition, increased nuchal translucency in karyotypically normal fetuses has been reported to predict additional categories of abnormal outcomes, such as cardiac and other structural anomalies. Further studies will, of course, be needed to verify and further quantify these risks. Midtrimester serum markers are not likely to further elucidate such nonchromosomal risks. Given abnormal first-trimester information, a patient might elect to act rather than wait for the second-trimester tests. These constitute some of the arguments for disclosing first-trimester information in a clinical setting. If, however, the combined first-trimester and midtrimester testing proves to be clearly superior to the other algorithms, then making it clinically available to patients with appropriate counseling would be justified. Prospective data are currently being gathered for the integrated first-trimester and midtrimester testing, both in the United States and Europe.

It is important to bear in mind that, of all the markers that have been prospectively and extensively tested in low-risk populations, first-trimester nuchal translucency is the most sensitive and specific and therefore constitutes the current “gold standard” for Down syndrome screening. In addition, there are distinctive advantages to first-trimester screening, such as early diagnosis and decision making regarding pregnancy treatment. Thus, until other promising algorithms have been extensively tested clinically, it would be premature to regard them as practical alternatives to established first-trimester markers in areas in which the latter are currently being used clinically.

The present study is unique. This is the only report in which available midtrimester urine, serum, and ultrasonographic scans have been compared directly in the same patients. This is crucial for the determination of the ideal marker combination for screening. In our previous studies, we looked at the diagnostic accuracy of a smaller combination of markers. There was, however, no competitive comparison of all the available markers, as was done in this study. The end result is a highly sensitive algorithm that requires a minimal number of simple markers.

In summary, this is the first report that has compared multiple urine, serum, and ultrasound markers in the same patient. The CMT had high Down syndrome sensitivities with low false-positive rates in a high-risk population. Several features of the algorithm suggest that it might be useful for low-risk population screening. Only 3 markers were used, which is desirable to minimize clinical variability and control costs. Maternal age did not contribute significantly to the algorithm, which suggests that good diagnostic performance could be achieved with the algorithm in a younger population. In a subset of patients who were <35 years old, this was confirmed. The number of cases in this subgroup was small, however. In addition, gross anomaly detection did not contribute significantly to the algorithm. The only ultrasound measurement required is nuchal thickness. This suggests that high diagnostic accuracy is still possible in areas or clinical sites with modest diagnostic ultrasound capabilities. Standard-

### Table II. Down syndrome sensitivity in the advanced maternal age* only group

<table>
<thead>
<tr>
<th>False-positive rate (%)</th>
<th>Sensitivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>69.2</td>
</tr>
<tr>
<td>0.6</td>
<td>76.9</td>
</tr>
<tr>
<td>0.8</td>
<td>92.3</td>
</tr>
</tbody>
</table>

Other measurements (serum AFP, hCG, uE₃, humerus length, gross anomaly) did not achieve statistical significance and were therefore excluded from the algorithm.

*Only indication for amniocentesis.

### Table III. Comparison of Down syndrome detection rates that were reported for first-trimester and midtrimester algorithm tests for Down syndrome detection

<table>
<thead>
<tr>
<th>Test/algorithm (reference)</th>
<th>1% FPR</th>
<th>5% FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuchal translucency¹</td>
<td>—</td>
<td>77</td>
</tr>
<tr>
<td>Nuchal translucency plus PAPP-A, free β-hCG²</td>
<td>72</td>
<td>85</td>
</tr>
<tr>
<td>Integrated test*²</td>
<td>85</td>
<td>94</td>
</tr>
<tr>
<td>CMT (overall group)</td>
<td>86</td>
<td>94</td>
</tr>
</tbody>
</table>

FPR, False-positive rate.
*First-trimester nuchal translucency and PAPP-A, midtrimester serum AFP, hCG, estriol, and inhibin-A.
ization, training, and periodic audits would need to be performed to achieve optimal diagnostic performance from nuchal-thickness measurement. The high diagnostic accuracy and need for only a urine specimen would make the test attractive to both patients and practitioners. The major limitation of the study is the relatively small sample size, particularly in the group <35 years old. Currently, a multicenter trial of the algorithm in women who undergo amniocentesis for maternal age is under way. Large population studies in women who are at average risk would be the next desirable and logical step. Before this, no general claims of clinical usefulness in an average-risk population can be made for the CMT.

REFERENCES


